International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

A Review: Protein Interaction & Behavior Assessment in Host Cells after Novel Drug Compound Administration using Systems Biology Approach

Yesha Modi[a], Harsh Shinde[b], Natasha Navet[c], Richa Arya[d], Nikita Sushil Kumar[e], Ved Mishra[f], Fariya Khan[g], Satyam Khanna[h], Ruchi Narula[i*], Prashant Agarwal[j]

[a,b,c,d,e] *Biotech Research Assistant, Biological Upliftment Academic Foundation, Kanpur, U.P. – India*

[f,g] *Computational Biology Research Head, Biological Upliftment Academic Foundation, Kanpur, U.P.-India*

[h,i] *Research Head, Rass Biosolution Pvt. Ltd., Kanpur, U.P.-India*

Abstract

To understand complex biological systems requires the integration of experimental and computational research; in other words systems biology approach. Computational biology, through via different software helps in exploration more than one gene expression at a time and also understanding the connectivity, Systems Biology provides a powerful foundation from which to address critical scientific questions head-on. The reviews in this Insight cover many different aspects of this energetic field, although all, in one way or another, illuminate the functioning of modular circuits, including their robustness, design and manipulation. Computational systems biology addresses questions fundamental to our understanding of life, yet progress here will lead to practical innovations in medicine, drug discovery and engineering, In this study we have evaluated the potentiality of Antifungal Aqueous extracts on Yeast cultures and scientifically proven the same using Cytoscape**.**

Keywords: Systems Biology, Proteomics, Protein Profiling, Cytoscape, Network, Nodes & Edges, Attributes, Annotation, Ontology, Gene Ontology etc.

---

* Corresponding author.

E-mail address: rankyicy@gmail.com.

1. Introduction

System biology is defining biochemical networks where biomolecules are the nodes and the molecular interactions between them are the edges [8]. Systems Biology is rapidly evolving field of computer science, mathematics, physics and biology. It endeavors to study, analyze and understand complex biological systems by taking a coordinated integrated systems view using computational methodologies [21**]**. Open and unsolved problems in biology range from understanding structure and dynamics of biological systems to prediction and inference in the complex systems. In the post genomic era, systems based approaches may provide a solution to such unsolved problems. It is believed that some answer to the question "what is life" may be obtained by taking a broader, and integrated biological view [24**]**.

Systems biology approaches can provide a solution to the key issue that is how to design lifesaving and cost-effective drugs so that the diseases can be cured and prevented. Pharmaceutical companies' view that systems based computational techniques will be highly useful in designing effective therapeutic drugs [9,20,23**]**. To increase the productivity of drug discovery one needs a far deeper understanding of the molecular mechanisms of diseases, taking into account the full biological context of the drug target and moving beyond individual genes and proteins [4,7,13**]**.

Expression profiling and large-scale proteomics have revolutionized biology by generating vast amounts of data about cell state. Genes with significant changes in expression have immediate and widespread interest as markers for diseases, stages of development, and a variety of other cellular phenotypes [1]. Genes with correlated expression changes over many conditions are likely to be involved in similar functions or cellular processes; these genes often also share DNA sequence elements, providing evidence that they are regulated by common transcription factors. Analytical methods such as gene expression clustering [12, 30], significance testing [19, 17, 27], and sequence motif identification [25] have been indispensable for enabling these discoveries and summarizing the data at each step.

For model organisms such as yeast, new technologies and data sets are making it possible to address these questions more directly than ever before. Systematic two-hybrid screens and co-immuno precipitation experiments are populating the public databases with thousands of protein-protein interactions and complexes [15, 32]. Other ongoing projects are defining large numbers of protein→DNA interactions [26], and protein microarrays are making it possible to map interactions between proteins and drugs, hormones, and other small molecules [33]. These molecular interactions provide a convenient framework for understanding changes in gene expression and for integrating a wide variety of global state measurements.

One basis of systemic approaches to biological processes is the knowledge generated from genome sequencing and large-scale genetic analyses revealing an already enormous scale of interactions on the level of nucleic acids [6, 10, 31]. A general understanding of the topology of genetic-interaction networks in yeast has a wider importance, because similar networks are expected to underlie the relationship between genotype and phenotype in higher eukaryotic species. Most of the insights into genetic interactions and networks come from studies using the yeast *Saccharomyces cerevisiae*, in which functional genomics tools allow systematic analyses [11].

1.1 Current Systems Biology Practices

Saccharomyces Genome Databases (SGD):

Saccharomyces Genome Databases (SGD) (www.yeastgenome.org) is a scientific database of the molecular biology and genetics of the yeast *Saccharomyces cereviseae*, has recently developed several new resources that allow the comparison and integration of information on a genome-wide scale, enabling the user not only to find detailed information about individual genes, but also to make connections across groups of genes with common features and across different species. The *Saccharomyces* Genome Database (SGD) was established to provide a convenient means for accessing the rapidly expanding knowledge available for the budding yeast *Saccharomyces cerevisiae*.

The completion of the *S.cerevisiae* genomic DNA sequence in 1996 provided the sequence of each of its genes and currently represents the only complete sequence of a eukaryotic genome. Systematic efforts to identify *S.cerevisiae* genes, describe their role within the cell's life cycle, and reveal their interactions with other gene products are now underway. Such experimental approaches are changing how basic biological research is conducted and are resulting in an explosion of information. The data in SGD are organized around the genome's sequence and its genes. SGD has as its primary goals the provision of information about the DNA sequence and its individual components, RNAs, encoded proteins and the structures and biological functions of any known gene products. An equally important goal of SGD has been to create tools which allow the user to easily retrieve and display these types of information. This has resulted in graphical interfaces which are geared towards biologists using the database, irrespective of their familiarity with yeast. By knowing a bit of sequence, a gene name, a function (e.g. enzymatic activity), or a map position, one can efficiently query for information about a gene. In addition, SGD serves as the *S.cerevisiae* community's repository for gene nomenclature [16].

1.2 Software for use

Cytoscape with its plug-ins:

Cytoscape is an open source bioinformatics software tool which is used for the visualizing molecular interaction networks and it integrating with gene expression. Cytoscape was developed at the Institute of System Biology in Seattle in 2002. It was made publically in July 2002. It is written in java and used in any java-based operating systems. There are many additional plugin available for network and molecular profiling analysis, new layouts, additional file format support and connection with databases and searching in large networks. This software downloads from website- www.cytoscape.org. Cytoscape is a project dedicated to building open-source network visualization and analysis software. Software "Core" provides basic functionality to layout and query the network and to visually integrate the network with state data. The Core is extensible through a plug-in architecture, allowing rapid development of additional computational analyses and features [5, 28). In present work we used the following four plug-ins-

BioNetBuilder: BioNetBuilder is an open-source client-server Cytoscape plug-in that offers a user-friendly interface to create biological networks integrated from several databases. Users can create networks for ~1500

organisms, including common model organisms and human. Currently supported databases include: DIP, BIND, Pro-links, KEGG, HPRD, The BioGrid, and GO, among others. The BioNetBuilder plug-in client is available as a Java Web start, providing a platform independent network interface to these public databases [3, 28].

jActiveModulus: Identification of Modules of Seed and Neighboring Genes Using the jActiveModule Method (jAM).The MT method results in a large number of statistically significant predictions, but some of the predictions may be artifacts of low or excessive connectivity. To address this concern, the jActiveModule method is implemented to determine modules with maximal proportions of the lowest *p* value genes. jActiveModule Plug-in enumerates those ids which are actively participating in one or more function after the incorporation of compound in the microbe has already taken place. The jActiveModule plug-in gives those ids which are point of interest of our result these ids specifically are the once which depict to be actively taking post in the functioning of the microbes [18].

BiNGO: The Biological Networks Gene Ontology tool (BiNGO) is an open-source Java tool to determine which Gene Ontology (GO) terms are significantly overrepresented in a set of genes. BiNGO can be used either on a list of genes, pasted as text, or interactively on sub graphs of biological networks visualized in Cytoscape. BiNGO maps the predominant functional themes of the tested gene set on the GO hierarchy, and takes advantage of Cytoscape's versatile visualization environment to produce an intuitive and customizable visual representation of the results. The main advantage of BiNGO over these tools is its interactive use on molecular interaction networks, e.g. protein interaction networks or transcriptional co-regulation networks, visualized in Cytoscape. Furthermore, BiNGO offers great flexibility in the use of ontologies and annotations. Besides the traditional GO ontologies, BiNGO also supports the use of GO Slim ontologies, as well as custom ontologies and annotations. Finally, the Cytoscape graphs produced by BiNGO can be viewed, laid out, modified and saved in various manners [2, 22].

MCODE: "Molecular Complex Detection" (MCODE), that detects densely connected regions in large protein-protein interaction networks that may represent molecular complexes. The method is based on vertex weighting by local neighborhood density and outward traversal from a locally dense seed protein to isolate the dense regions according to given parameters. The algorithm has the advantage over other graph clustering methods of having a directed mode that allows fine-tuning of clusters of interest without considering the rest of the network and allows examination of cluster interconnectivity, which is relevant for protein networks. Protein interaction and complex information from the yeast *Saccharomyces cerevisiae* was used for evaluation [14].

2. Conclusion and Contribution to the society

Preventive medicine will follow as disease- perturbed networks can be used to identify drug targets—first for therapy and later for prevention. Pharmacological intervention will focus on preventing disease-mediated transitions, as well as reversing or terminating those that have occurred. This will require building a fundamental understanding of the systems biology that underlies normal biological and pathological processes, and the development of new technologies that will be required to achieve this goal.

Computational biology and bioinformatics approaches have the potential to completely change the way drugs are discovered and designed. Computational methods like classification and network-based algorithms can be used to understand the mode of action and the efficacy of a given compound and to help elucidating the pathophysiology of a disease. But these computational tools, in our opinion, may also be used in a different and innovative way to promote a change of paradigm in how drugs are designed. In the pharmacological industry there has already been a shift from symptomatic oriented drugs that can relieve the symptoms but not the cause of the disease to pathology-based drugs whose targets are the genes and proteins involved in the etiology of the disease. Drugs targeting the affected pathway have thus the potential to become therapeutic. A network approach to drug design would examine the effect of drugs in the context of a network of relevant protein-protein, regulatory and metabolic interactions. The end result would be the development of a drug that would hit multiple targets selected in such a way as to decrease network integrity and so completely disrupt the functioning of the network. The screening of a compound to quickly identify the proteins it interacts with gives us all the necessary tools to identify and repair the disregulated biological pathway causing the disease, much as an engineer would do to restore a malfunctioning electronic circuit. Cytoscape is a general-purpose, open-source software environment for the large scale integration of molecular interaction network data. Dynamic states on molecules and molecular interactions are handled as attributes on nodes and edges, whereas static hierarchical data, such as protein-functional ontologies, are supported by use of annotations. The Cytoscape Core handles basic features such as network layout and mapping of data attributes to visual display properties.

In the present work for the identification of mode of action of natural compound comparative expression proteomics approach was applied. The complex genetic landscape and interaction networks were analyzed to significantly identify the response of the compound. This approach gives protein based validation on a specific natural compound as a potent therapeutic agent. Computational framework assessment on the data gives percentage prediction on the natural compound’s potency as a therapeutic agent.

The application of systems biology to medical practice is the future of medicine. Its realization will see drug discovery and the design of multiple drug therapies and therapeutic gene circuits being pursued just as occurs now with modern, complex engineering products. Although the road ahead is long and winding, it leads to a future where biology and medicine are transformed into precision engineering. In the near future, the most pressing task is to investigate our identified sub networks in the laboratory. Because large interaction networks are suspected to contain many false-positives, an initial experiment would be to verify that the interactions in each sub network are reproducible and present under the subnet’s particular set of conditions. Routine network screening has to be performed to define novel modes of regulation, to identify evolutionarily conserved pathways, or to interrogate regulatory circuits responding to the entire spectrum of drugs and human diseases.

References

[1] Altman, R. B., and Raychaudhuri, S. (2001). Whole-genome expression analysis: challenges beyond clustering, Curr Opin Struct Biol 11, 340-7.

[2] Ashburner M., Ball C. A., Blake J. A. and Botstein D., (2000), The Gene Ontology: tool for the unification of biology, Nature Genetics, 25, 25-29.

[3] Avila-Campillo I., Drew K., Lin J., Reiss D. J. and Bonneau R., (2007), BioNetBuilder: automatic integration of biological networks; Bioinformatics1:23(3): 392-3, EPub, PMID: 171385.

[4] Apic G., Ignjatovic T. and Boyer S., (2005), Illuminating drug discovery with biological pathways. FEBS Lett. 579, 1872-1877.

[5] Bell G. W. and Lewitter F. (2006), Visualizing networks; Meth Enzymol. 411:408-21. Doi: 10.1016|50076-6879(06)11022-8 PMID.16939803.

[6] Boone C., Howard B. and Brenda J. Andrews, (2007), Exploring genetic interactions and networks with yeast. Nat. Rev. Genet. 8, 437-449.

[7] Butcher E. C., Berg E. L., and Kunkel E. J., (2004), System biology in drug discovery. Nat Biotechnol, 22:1253-1259.

[8] CajaMurcia F., (2006), Red Espanola de Biología de Sistemas, Systems biology, 14(1):19–31.

[9] Cho C. R., Mark Labow, Mischa Reinhardt, Jan van Oostrum and Manuel C. Peitsch (2006), The application of systems biology to drug discovery. Current Opinion in Chemical Biology, 10(4):294–302.

[10] Davierwala A. P., Haynes J., Li Z., Brost R. L., Robinson M. D., and Yu L., (2005), The synthetic genetic interaction spectrum of essential genes. Nat. Genet. 37, 1147-1152.

[11] Dolinski K. and Botstein D., (2005), Changing perspectives in yeast research nearly a decade after the genome sequence. Genome Res., 15, 1611-1619.

[12] Eisen M. B., Spellman P. T., Brown P. O., and Botstein D. (1998). Cluster analysis and display of genome-wide expression patterns, Proc Natl Acad Sci U S A 95, 14863-8.

[13] Fishman M. C. and Porter J. A., (2005), A new grammar for drug discovery. Nature, 437:491-493.

[14] Gary D. B. and Christopher W. V. H., (2003), An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics, 4:2 doi 10, 1186|1471-2105-4-2.

[15] Gavin A.C., Bösche M., Krause R., Grandi P., and Marzioch M. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes, Nature 415, 141-147.

[16] Goffeau A., Barrell B. G., Bussey H., Davis R. W., Dujon B., Feldmann H., Galibert F., Hoheisel J. D., Jacq C., Johnston M., Louis E. J., Mewes H. W., Murakami Y., Philippsen P., Tettelin H., and Oliver S. G., (1996), Life with 6000 genes. Science, 274, 546.

[17] Ideker T., Thorsson V., Siegel A., and Hood L. (2000). Testing for differentially-expressed genes by maximum likelihood analysis of microarray data, J Comput Biol 7, 805-817.

[18] Ideker T., Ozier O., Schwikowski B., and Siegel A. F., (2002), Discovering regulatory and signaling circuits in molecular interaction networks. Bioinformatics. 18 suppl 1:5233-40; PMID: 12169552.

[19] Kerr, M. K., and Churchill, G. A. (2001). Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments, Proc Natl Acad Sci U S A 98, 8961-5.

[20] Kumar N., Hendriks BS, Janes KA, Graaf D, Lauffenburger DA (2006), Applying computational modeling to drug discovery and development. Drug Discovery Today, 11(1718): 806–811.

[21] Lodhi H. and Muggleton S., (2004), Modeling metabolic pathways using stochastic logic programs based Ensemble methods. Second International Conference on Computational Methods in System Biology (CMSB04), LNCS, pages 119–133.

[22] Maere Steven, Heymans K. and Kuiper M., (2005). BiNGO: a cytoscape plugin to access overrepresentation of gene ontology categories in biological networks. Bioinformatics 21, 3448-3449, doi: 10.1093|bioinformatics|bti551.

[23] Materi W. and Wishart S., (2007), Computational systems biology in drug discovery and development: methods and applications. Drug Discovery Today, 12(78): 295–303.

[24] Noble D., (2006), The Music of Life, Oxford University Press, pages 263-274.

[25] Pilpel, Y., Sudarsanam, P., and Church, G. M. (2001). Identifying regulatory networks by combinatorial analysis of promoter elements, Nat Genet 29, 153-9.

[26] Ren B., Robert F., Wyrick J. J., Aparicio O., Jennings E. G., Simon I., Zeitlinger J., Schreiber J., Hannett N. and Kanin E., (2000). Genome-wide location and function of DNA binding proteins, Science 290, 2306-9.

[27] Rocke, D. M., and Durbin, B. (2001). A model for measurement error for gene expression arrays, J Comput Biol 8, 557-69.

[28] Shannon P., Markiel A., and Ozier O., (2003), Cytoscape: a software invironment for integrated models of biomolecular interaction networks; Genome Res. 13(11):2498-504. Doi: 10.1101|gr|239303. PMC 403769, PMID 1457658.

[29] Shannon P. T., David J Reiss**,** Richard Bonneau and Nitin S. Baliga, (2006), The Gaggle: an open-source software system for integrating bioinformatics software and data sources, BMC Bioinformatics, 7, 176.

[30] Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S., and Golub, T. R. (1999). Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation, Proc Natl Acad Sci U S A 96, 2907-2912.

[31] Tong A. H. Y., Lesage G, Bader GD, Ding H, Xu H, Xin X, and Young J., (2004), Global mapping ofthe yeast genetic interaction network. Science, 303, 808-813.

[32] Uetz, P., Giot L., Cagney G., Mansfield T. A., Judson R. S., Knight J. R., Lockshon D., Narayan V., Srinivasan M., Pochart P., Qureshi-Emili A., Li Y., Godwin B., Conover D., Kalbfleisch T., Vijayadamodar G., Yang M., Johnston M., Fields S. and Rothberg J. M., (2000), A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. J. M. Nature, 403, 623-627.

[33] Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S. and Houfek, T., (2001). Global analysis of protein activities using proteome chips, Science 293, 2101-5.